Stability criteria for yet another class of multidimensional distributed systems

Leonidas Georgiadis, Wojciec Szpankowski

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Stability Criteria for Yet Another Class of Multidimensional Distributed Systems

Leonidas GEORGIADIS
Wojciech SZPANKOWSKI

N° 1996
Juillet 1993

PROGRAMME 2

Calcul symbolique,
programmation
et génie logiciel

Rapport de recherche

1993

STABILITY CRITERIA FOR YET ANOTHER CLASS OF MULTIDIMENSIONAL DISTRIBUTED SYSTEMS

September 1991

Revised: June 30, 1993

Leonidas Georgiadis
IBM T. J. Watson Research Center
P.O. Box 704
Yorktown Heights, NY 10598
U.S.A.
leonid@watson.ibm.com

Wojciech Szpankowski*
Department of Computer Science
Purdue University
W. Lafayette, IN 47907
U.S.A.
szpankow@margaux.inria.fr

Abstract

We present sufficient and necessary conditions for stability of token passing rings with time-limited discipline and more generally, a class of monotonic and contractive service disciplines. In general, establishing stability for multidimensional distributed systems is a difficult problem. The standard Lyapunov test function method often fails when applied to such systems (e.g., token passing rings, ALOHA-type systems, rings with spatial reuse, couple-processors system, etc.). In our recent work on this topic we establish a useful alternative approach that turns out to be successful for deriving stability conditions for several distributed systems. In the current paper, we show that our approach can be extended so that it can be applied to the system under consideration.

CRITERES DE STABILITE POUR UNE NOUVELLE CLASS DE SYSTEMES DISTRIBUE MULTIDIMENSIONNELS

Résumé

Nous présentons des conditions nécessaires et suffisantes pour la stabilité d'anneaux à jeton avec discipline de service sous contraintes temporelles. En général, la stabilité de systèmes multidimensionnels est un problème difficile. Les tests à la Lyapunov y sont souvent voués à l'echec (anneau à jeton, systèmes de type ALOHA, anneau avec réutilisation spatiale, couplage de processeurs, etc.). Dans un de nos récents travaux sur ce sujet nous établissons une méthode alternative qui s'avère efficace pour établir la stabilité de nombreux systèmes. Dans ce papier, nous montrons que notre approche peut être étendue à une classe plus large de systèmes.

*This research was supported in part by the NSF grants NCR-9206315, CCR-9201078, and INT-8912631, by AFOSR grant 90-0107, and in part by NATO Collaborative Grant 0057/89. This paper was revised while the author was visiting INRIA, Rocquencourt, France, and he wishes to thank INRIA (projects ALGO, MEVAL and REFLECS) for a generous support.

1. INTRODUCTION

Distributed multiqueue systems which share a single scarce resource (i.e., server) such as a communication channel or a processor, have received a considerable amount of attention in the recent literature. Of special interest are token passing rings (cf. [5], [16], [17]) due to a number of reasons. For example, it is known that obtaining the distribution of the number of messages queued in each station is a formidable open problem, as is the problem of obtaining the waiting time distribution. The stability condition for the token passing ring was *heuristically* predicted by Kuehn [5] in 1979, and then reproduced with some minor changes in many other papers (cf. [4]). Recently, in our paper [3] we have established rigorously stability conditions for the ℓ-limited token passing rings, and indicated that the technique can be extended to other disciplines We complete this study in the current paper. Specifically, we apply our methodology to derive stability criteria for time-limited token passing rings introduced recently by Leung and Eisenberg [6], [7] (cf. Proposition 1). In such a system each station transmits messages for at most an amount of time τ. If the transmission time exceeds τ, the station completes the transmission of the message in progress and sends the token to the next station (the so called nonpreemptive time limited discipline). While we study this system as an important application, the technique can be applied almost without modification to a class of monotonic and contractive policies (cf Proposition 2).

Our approach to the stability of token passing rings follows the idea discussed in our paper [3], and differs significantly from the standard methodology of the Lyapunov test function (cf. [12], [18], [14]). (For other than test function approaches see also [2], [11] [14, 15].) It resembles the general idea of Malyshev's *faces* and *induced Markov chains* [10]. Our method is based on a simple idea of stochastic dominance technique, and application of Loynes [9] stability criteria for an isolated queue. We note that this approach is *not* restricted to token passing rings, and stability of several other distributed systems can be assessed by this methodology (cf. [14, 15]).

We now summarize our main results. We shall analyze the token passing ring with Poisson arrivals with parameter λ_i for the ith station, general distribution of service times $\{S_i^k\}_{k=1}^{\infty}$ and switchover times $\{U_i^k\}_{k=1}^{\infty}$. We consider a gated version of τ-limited token ring. Define $\widetilde{L}_i = \min\{k : \sum_{j=1}^{k} S_i^j \geq \tau_i\}$ and $\ell_i = E\widetilde{L}_i$. Clearly, ℓ_i is the average of the maximum number of customers served during τ_i.

Proposition 1. *Consider a token passing ring consisting of M stations with τ-limited service schedule for the ith station, and Poisson arrivals. Then the system is stable if and* *only if* $\sum_{j=1}^{M} \rho_j < 1$ *and*

$$\lambda_j < \frac{\ell_j}{u_0}(1 - \rho_0) \quad \text{for all} \quad j \in \mathcal{M} = \{1, \dots M\} \ ,$$

where $u_0 = \sum_{i=1}^{M} EU_i$ *is the average total switchover time, and* $\rho_0 = \sum_{i=1}^{M} \rho_i$ *with* $\rho_i = \lambda_i s_i$ *and* $s_i = ES_i$ *being the average service time at the ith station.* ■

It will be seen that the method of the proof can be applied virtually unchanged to the following class of monotonic and contractive policies. We assume that the number of customers served from queue i when there are n queued messages at the instant of token arrival to queue i is $f_i(n, \mathbf{X})$, where $\mathbf{X}$ is a possibly random quantity that depends on the policy. We assume that $f_i(\cdot, \mathbf{X})$ is a *nondecreasing* function of the number of customers in the ith queue. In addition, the following relation holds

$$f_i(n_1, \mathbf{X}) - f_i(n_2, \mathbf{X}) \leq n_1 - n_2 \qquad if \qquad n_1 > n_2 \ .$$

At the kth token arrival to queue i the quantity $\mathbf{X}$ takes the value $\mathbf{A}_k$. The random variables $\{\mathbf{A}_k\}_{k=1}^{\infty}$ are i.i.d, are independent of the past arrival times, service times and switchover times. Note that $f_i(n, \mathbf{A}_k)$ tends to a random variable, $f_i^*(\mathbf{A}_k)$, when $n \to \infty$. We assume that $f_i^*(\mathbf{A}_k)$ is finite and

$$\ell_i \stackrel{\text{def}}{=} Ef_i^*(\mathbf{A}_k) < \infty \ .$$

We will see in the next section that the time-limited token ring policy falls within the class just described. Other special cases are the ℓ-limited gated discipline and Bernoulli gated disciplines. In the last descipline, $f_i(n, X) = \min\{n, X\}$, where X is a geometrically distributed random variable. The following result follows directly from the arguments that will be presented in this paper.

Proposition 2. *Consider a token passing ring that employs a monotonic and contractive service discipline described above. Then, the system is stable if and only if*

$$\lambda_i < \frac{\ell_i}{u_0}(1 - \rho_0) \qquad \text{for all} \quad i \in \mathcal{M}$$

where $\rho_0 < 1$. ■

One policy that does not fall in the previously described category is the *preemptive* τ-*limited* token ring. In such as system the token interrupts his service immediately after the time limit τ expires and continues servicing the same customer in the next round. While the formal arguments of our methodology can be applied for this system, there are technical difficulties that need to be overcome for a complete proof. We believe that these technical arguments can be provided, but we do not have a complete proof yet, and therefore we express the following conjecture.

Conjecture. *The preemptive τ-limited token ring is stable if and only if*

$$\lambda_j < \frac{\tau_j}{s_j u_0}(1 - \rho_0) \quad \text{for all} \quad j \in \mathcal{M} = \{1, \dots M\} \ ,$$

and $\rho_0 < 1$. ∎

The paper is organized as follows. In the next section we present some preliminary results In particular, we find Markovian representations of the system, prove a crucial stochastic dominance relationship, and establish some Wald's type formulas. Finally, in Section 3 we present our main construction that leads to the proof of the Proposition 1. Since the proof of Proposition 2 is along the lines of the proof of Proposition 1, we shall concentrate only on establishing Proposition 1, leaving the details of the proof of Proposition 2 to the reader.

2. PRELIMINARY RESULTS

In this section we present several results that are required to establish our main finding regarding the stability of the token passing ring. These results are of their own interests, and can be used to obtain some estimates for the performance evaluation of the system. In the sequel, we list our main assumptions, describe the Markovian character of an imbedded queueing process, show two simple Wald's type identities, and finally establish a stochastic dominance relationship.

We start with a precise definition of our stochastic model. We shall adopt the following assumptions.

(A1) There are M stations (queues) on a loop, each having infinite capacity buffer.

(A2) Maximum time customers are served during the token visit at a queue is limited to $\tau_i < \infty$ units of time. Only customers that are present at the instant of token arrival can be served. Moreover, we have nonpreemptive discipline, that is, the customer that is in the server when the time limit τ_i is reached, is served to completion before the token moves to the next queue.

(A3) Arrival proces A_i^t, $t \in [0, \infty)$ to the ith queue is a Poisson process with parameter $\lambda_i > 0$. Here, A_i^t is the number of arrivals at queue i up to time t. The arrival process at a queue is independent of the arrival processes to other queues.

(A4) Service time process $\{S_i^k\}_{k=1}^{\infty}$ at queue i is i.i.d. with $s_i = ES_i^1 > 0$. The service time process at a queue is independent of the arrival processes at all queues and independent of the service time processes at other queues.

(A5) The switchover times between i and $i+1$ mod M queue, $\{U_i^k\}_{k=1}^{\infty}$, are i.i.d., independent of the switchover times $\{U_j^k\}_{k=1}^{\infty}$ for $j \neq i$, and independent of the arrival and the service time processes. The average total switchover time is defined as $u_0 = \sum_{i=1}^{M} EU_i^1$. To avoid unnecessary complications we assume that $P(U_i^n > 0) = 1, \quad i = 1, \cdots, M$.

Now we are ready to present a Markovian description of the system. We need a little bit of notation. By (A1), the token visits stations in a cyclic order. Let n denote the nth visit of the token to any queue. Then, $k_n = \lfloor (n-1)/M \rfloor + 1$ denotes the cycle number in which the nth visits occurs (we start counting cycles from one and assume that the token starts from queue 1). Note that the queue visited at the nth visit is just $J_n = n - M(k_n - 1)$. Let also $\{T_n\}_{n=1}^{\infty}$ be the time instant of the nth visit of the token to any queue. Define an M-dimensional process $\tilde{\mathbf{N}}^n = (\tilde{N}_1^n, \ldots, \tilde{N}_M^n), \quad n = 1, 2, \cdots,$, where $\tilde{N}_i^n$ as the number of customers in queue i at time T_n. In addition, by $\mathcal{N}_i^n$ we mean the total number of customers served from queue i up to time T_n. Theorem 1 below shows that $\tilde{\mathbf{N}}^n$ is a Markov chain. Since its proof is along the lines of our Theorem 1 from [3], we omit it here.

Theorem 1. *The process $\tilde{\mathbf{N}}^n$ is a nonhomogeneous Markov chain.* ■

Remark: It should be noted that the assumption that the service discipline is nonpreemptive (see assumption (A2)) is crucial for Theorem 1 to hold. Assume, for example, that preemptions were allowed so that a server could interrupt the service of the customer as soon as the limit τ was reached. Upon the next arrival of the token to the queue, the server could either complete the remaining service time or restart the service time of the interrupted customer. In both cases, the number of customers in queue i at time T_n will depend in general on the service time $S_i^{\mathcal{N}_i^n+1}$ and the process of queue lengths will not be Markov.

There are other Markovian descriptions of the system. For example, define $N_j^n(i)$ to be the number of customers at queue j when the token visits queue i for the nth time. Then, the process $\mathbf{N}^n(i) = (N_1^n(i), \ldots, N_M^n(i))$ can be deduced from $\tilde{\mathbf{N}}^n$ since $\mathbf{N}^n(i) = \tilde{\mathbf{N}}^{(n-1)M+i}$ and therefore it is a Markov chain. It is not difficult to verify also that,

Corollary 2. *The process $\mathbf{N}^n(i)$ of the queue lengths registered by the token when it visits (reference) queue i, is a homogeneous, irreducible and aperiodic Markov chain.* ■

The fact that under assumption (A2) the service times of the customers at queue i at instant T_n are i.i.d. independent of the queue size at time T_n, permits us to consider a new model of the system which is stochastically equivalent to the original one and has the advantage that under this new model, many of the arguments that follow become simpler. Specifically, in the new system assumptions (A1)-(A3) and (A5) are the same, while assumption (A4) is replaced with

(A4') Service times are assigned to the customers at queue i upon beginning of service as follows. We consider a doubly infinite sequence of i.i.d random variables $\{S_i^{n,k}\}_{n,k=1}^{\infty}$ with $s_i = ES_i^{n,k} > 0$. The customers that are served during the nth arrival of the token to queue i are assigned the service times $S_i^{n,1}, S_i^{n,2}, \cdots$. The sequence $\{S_i^{n,k}\}_{n,k=1}^{\infty}$ is independent of the sequence $\{S_j^{n,k}\}_{n,k=1}^{\infty}$ for $i \neq j$, and independent of the interarrival processes to the queues.

Next, we need some relationships between the average number of customers served per token visit L_i^n and the average cycle time C_i^n. The former quantity is defined as follows. Let $\widetilde{L}_i^n = \min\{k : \sum_{j=1}^{k} S_i^{n,j} \geq \tau_i\}$. Then $L_i^n = \min\{\widetilde{L}_i^n, N_i^n\}$. The latter quantity is the length of time between the nth and $n+1$st visits of the token to the reference queue i. By EL_i and EC we denote the limiting averages of L_i^n and C_i^n. It turns out that the relations holding for the l-limited case, continue to hold under the more general policies we consider here. Specifically, we have the following result.

Theorem 3. *Let the Markov chain* $\mathbf{N}^n(i)$ *be positive recurrent (ergodic) for some* $i \in \mathcal{M}$. *Then,* $\mathbf{N}^n(j)$ *is ergodic for all* $j \in \mathcal{M}$, *and* $\rho_0 = \sum_{j=1}^{M} \rho_j < 1$. *In addition,*

$$EL_j = \lambda_j EC, \qquad j \in \mathcal{M} \tag{1}$$

$$EC = \frac{u_0}{1 - \sum_{j=1}^{M} \rho_j}, \tag{2}$$

where u_0 *is the total average switchover time (cf. assumption (A5)) and* $\rho_j = \lambda_j s_j$ *is the utilization coefficient for the ith queue.*

Proof. We need only minor modifications compared to the proof of our Theorem 3 in [3]. We present a brief sketch of the proof. Without loss of generality, let $i = 1$. Define

$$K^{n+1} = \min\{m > K^n : \mathbf{N}^m(1) = \mathbf{0}\},$$

and $R^n = K^{n+1} - K^n$. We also write $R = R^1$. Due to the ergodicity of $\mathbf{N}^n(1)$ we have $ER < \infty$. Observe that for $j \in \mathcal{M}$, the process $\mathbf{N}^n(j)$ is regenerative with respect to R^n.

Therefore (cf. [1]), almost surely

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} L_j^k}{n} = \frac{E\left(\sum_{k=1}^{R} L_j^k\right)}{ER} = EL_j \quad , \quad \lim_{n\to\infty} \frac{\sum_{k=1}^{n} C_1^k}{n} = \frac{E\left(\sum_{k=1}^{R} C_1^k\right)}{ER} = EC_1 \ . \quad (3)$$

Let now $\tilde{L}_i^n = \min\{k : \sum_{j=1}^{k} S_i^{n,j} \geq \tau_i\}$. Clearly, $\tilde{L}_i^n$, $n = 1, 2, \cdots$ are i.i.d. random variables and since $s_1 > 0$, it is well known from renewal theory that $E\tilde{L}_1^n = \ell_1 < \infty$. Observe also that the event $\{R \leq k\}$ is independent of the sequence $\tilde{L}_1^n$, $n > k$ and therefore, $E\left(\sum_{k=1}^{R} \tilde{L}_1^k\right) = \ell_1 ER < \infty$. Note that in the interval $[0, \sum_{k=1}^{R} C_1^k)$ *all the arriving customers from all queues must be served.* If $\mathcal{A}_j$ is the number of arrivals to queue j in the interval $[0, \sum_{k=1}^{R} C_1^k)$, then $E\mathcal{A}_j = E\left(\sum_{k=1}^{R} L_j^k\right)$, and due to the Poisson assumption (A3) we also have

$$\lambda_j E\left(\sum_{k=1}^{R} C_1^k\right) = E\left(\sum_{k=1}^{R} L_j^k\right) \leq E\left(\sum_{k=1}^{R} \tilde{L}_1^k\right) < \infty \qquad (4)$$

The above and (3) lead to $EL_j = \lambda_j EC_1$, which completes the proof of (1). The proof of (2) is similar. ∎

The next result is one of the key elements of our stability analysis. We first give a brief overview of our approach, avoiding techinical details. In the process of estimating stability we need to build several dominant systems of the original token passing ring. For example, when we study stability of an isolated station, say the jth one, we partition all other stations into a class $\mathcal{S}$ of *nonpersistent* queues and a class $\mathcal{U}$ of *persistent* queues (saturated queues). A nonpersistent queue serves customers in the normal way as in the original token passing ring. A persistent queue, however, always sends the *maximum* allowable number of customers, that is, $\tilde{L}_i$ for $i \in \mathcal{U}$, by sending if necessary dummy customers. A question is whether such a new system dominates the original token passing ring in some sense. If the answer is *yes*, then by proving stability of the dominant system we establish stability of the original token passing ring. In passing, we note that persistent and nonpersistent queues are equivalent to "faces" and "induced Markov chains" in the Malyshev terminology ([10]).

We state the next result in a general form, since it is needed to prove Proposition 2. Specifically, in the terminology of [8], we consider the class of monotonic, contractive policies. This amounts to replacing assumption (A2) with the following more general one.

(A2') Let $\mathbf{A}$ denote a sequence of real numbers $\{a_1, a_2, \cdots\}$. Let $f_i(m, \mathbf{A})$ be the number of customers served from queue i when there are m queued messages at the instant of the nth token arrival at queue i and $\{S_i^{n,1}, S_i^{n,2}, \cdots\} = \mathbf{A}$. We assume that for fixed

$\mathbf{A}$, $f_i(m, \mathbf{A})$ is a nondecreasing function of m. In addition, for a fixed $\mathbf{A}$, the following relation holds

$$f_i(m_1, \mathbf{A}) - f_i(m_2, \mathbf{A}) \leq m_1 - m_2 \qquad if \qquad m_1 > m_2 \ . \tag{5}$$

For the case of τ-limited policy we have that

$$f_i(m, \{S_i^{n,1},\ S_i^{n,2}, \cdots\}) = \min\{m, \tilde{L}_i^n\}.$$

Now we are ready to formulate our result. Consider two token passing rings, say θ and Θ. Both satisfy assumptions (A1)-(A5) with (A2) replaced by the weaker assumption (A2'). The system θ represents our original token passing ring. The system Θ differs only in the switchover times, namely, we assume that the switchover time for Θ is replaced by $\{\Delta_i^k + U_i^k\}_{k=1}^{\infty}$ for $i = 1, \cdots, M$. We assume that for every $i \in \mathcal{M}$ and every $k \geq 0$ we have $\Delta_i^k \geq 0$. We make the following assumption for the process Δ_i^k.

(A6) The random variable Δ_i^k is independent of the service times, switchover times and the Poisson increments of the arrival processes to all stations after time $T_{M(k-1)+(i+1)} - U_i^k$ (see Fig.1).

Theorem 4. *Let $\tilde{\mathbf{N}}^n(\theta)$ and $\tilde{\mathbf{N}}^n(\Theta)$ denote the queue lengths in both systems. Then, under the above assumptions, and under the condition that the token starts from the same queue, say queue number one, and with the same number of initial customers in both systems, the following holds*

$$\tilde{\mathbf{N}}^n(\theta) \leq_{st} \tilde{\mathbf{N}}^n(\Theta), \tag{6}$$

where $\leq_{st}$ means stochastically smaller.

Proof. The proof is along the lines of Theorem 4 our paper [3]. For completeness in the presentation, we provide some details. To avoid cumbersome notation we present the proof only for $M = 2$ users.

We define some new variables. For a system θ let T_n^θ and D_n^θ denote the instances of the nth visit and the nth token departure from any queue respectively. As before, J_n^θ denotes the queue number visited at the nth visit of the token. Finally, $L_i^n(\theta)$ as before denotes the number of customers served from queue i at the nth visit of the token. Clearly, for our two station system $L_1^n(\theta) = 0$ for n even, and $L_2^n(\theta) = 0$ for n odd. In a similar manner we define respective quantities in the Θ system.

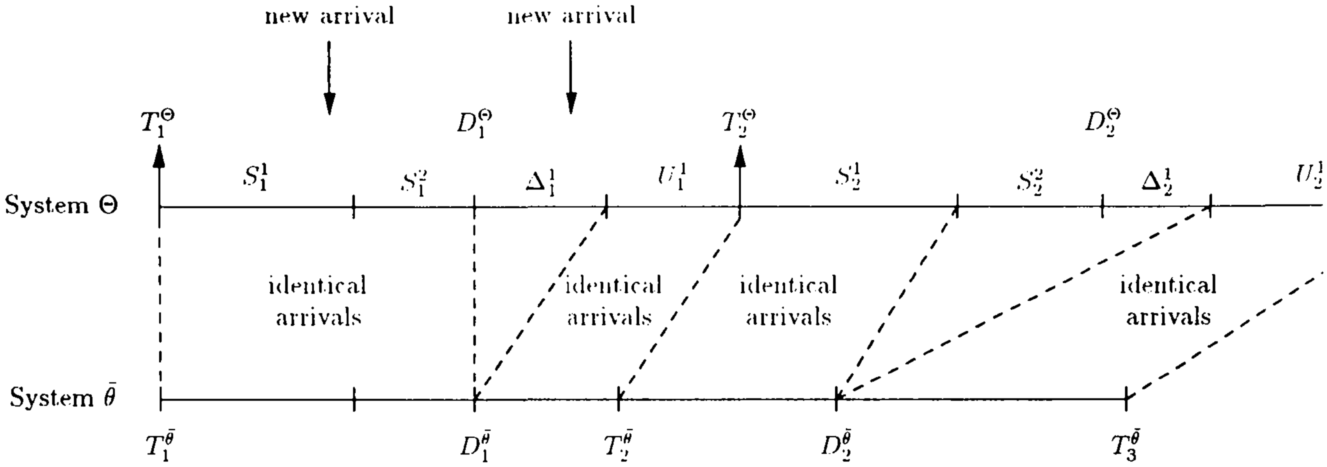

Figure 1: Illustration to the proof of Theorem 4

We will construct from the system Θ a token passing ring $\bar{\theta}$, which is stochastically equivalent to the system θ and for which we have that

$$\tilde{\mathbf{N}}^n(\bar{\theta}) \leq \tilde{\mathbf{N}}^n(\Theta) \,. \tag{7}$$

Figure 1 should help to understand our construction. Assume $\tilde{N}_i^1(\bar{\theta}) = \tilde{N}_i^1(\Theta)$ for $i = 1, 2$. The service times in system $\bar{\theta}$ are assigned from the same sequences $S_i^{n,k}$ as in Θ (according to assumption (A4').) Also, the same functions $f_i(m, \mathbf{A})$, $i = 1, 2$ are used in both systems. Therefore, the decision to switch to queue 2 will occur at the same time, namely $D_1^{\bar{\theta}} = D_1^\Theta$. The switchover time for $\bar{\theta}$ becomes now U_1^1, and of course $T_2^{\bar{\theta}} \leq T_2^\Theta$ since $\Delta_1^1 \geq 0$ (see Fig. 1).

The arrivals in the system $\bar{\theta}$ in $[D_1^{\bar{\theta}}, T_2^{\bar{\theta}})$ *are now assumed to be identical* to the arrivals in $[D_1^\Theta + \Delta_1^1, T_2^\Theta)$ in Θ system. Therefore clearly $\tilde{N}_i^2(\bar{\theta}) \leq \tilde{N}_i^2(\Theta)$ for $i = 1, 2$. The arrivals to system $\bar{\theta}$ in $[T_2^{\bar{\theta}}\ ,\ T_2^{\bar{\theta}} + S_2^1 + \cdots + S_2^{L_2^1(\bar{\theta})})$ are taken to be identical to the arrivals in $[T_2^\Theta\ ,\ T_2^\Theta + S_2^1 + \cdots + S_2^{L_2^1(\bar{\theta})})$. Note that this can be done since by (A2') $L_2^1(\bar{\theta}) \leq L_2^1(\Theta)$. Observe also that $T_2^{\bar{\theta}} + S_2^1 + \cdots + S_2^{L_2^1(\bar{\theta})} = D_2^{\bar{\theta}}$ (Fig. 1).

To complete the description of the system $\bar{\theta}$ we have to specify the arrivals in $[D_2^{\bar{\theta}}\ ,\ D_2^{\bar{\theta}} + U_2^1)$. These are taken to be exactly the arrivals in $[D_2^\Theta + \Delta_2^1\ ,\ D_2^\Theta + \Delta_2^1 + U_2^1)$ in the dominant system Θ (see Fig. 1). Note from the construction that

$$\tilde{N}_1^3(\bar{\theta}) = \tilde{N}_1^2(\bar{\theta}) + A_{[T_2^{\bar{\theta}}, T_3^{\bar{\theta}}]} \leq \tilde{N}_1^2(\Theta) + A_{[T_2^\Theta, T_3^\Theta]} = \tilde{N}_1^3(\Theta)$$

and also by (5),

$$\tilde{N}_2^3(\bar{\theta}) = \tilde{N}_2^2(\bar{\theta}) - L_2^2(\bar{\theta}) \leq \tilde{N}_2^2(\Theta) - L_2^2(\Theta) = \tilde{N}_2^3(\Theta) \ .$$

We can now repeat exactly the same procedure to construct $\bar{\theta}$ in the interval $[T_n^{\bar{\theta}}, T_{n+1}^{\bar{\theta}})$, $n \geq 3$, in the same manner as it was constructed in the interval $[T_2, T_3)$. By construction the service times and switchover times of system $\bar{\theta}$ are identically distributed to the corresponding variables of system θ and are independent of the interarrival process. In addition, assumption (A6) and the fact that the servicing policy is nonanticipative assures that the times $T_{n+1}^{\Theta} - U_{J_n}^{k_n}$ are stopping times for the Poisson arrival processes to all stations. The independence of the increments of the Poisson process implies now, that the constructed interarrival process in system $\bar{\theta}$ is Poisson with rate λ_i for queue i. Moreover, by construction (7) holds. Since $\bar{\theta}$ is stochastically equivalent to θ, we have that the distribution of $\tilde{\mathbf{N}}^n(\theta)$ is identical to the distribution of $\tilde{\mathbf{N}}^n(\bar{\theta})$. This completes the proof of Theorem 4. ■

3. MAIN RESULTS

Our approach in this section is based on three simple observations. *At first*, we note that a multidimensional process is stable if and only if its components are stable [14]. More precisely, if $\mathbf{N}^n = (N_1^n, \ldots, N_M^n)$ is a stochastic process – not necessary a Markov chain – then say the process is stable if the distribution of $\mathbf{N}^n$ as $n \to \infty$ *exists* and the distribution is *honest* (cf. [9, 14, 19]).

Secondly, to obtain stability conditions for a single isolated station in the token passing ring, we apply the technique of Loynes [9] who proved that a single $G|G|1$ queue is stable if the input rate is smaller than the average service time provided that service times and interarrival times are jointly stationary and ergodic. To verify a technical stationarity condition in Loynes' criteria we apply the stochastic dominance result of Theorem 4. More precisely, we partition the set of queues, $\mathcal{M}$, into a set $\mathcal{S}$ of nonpersistent queues and into a set $\mathcal{U}$ of persistent queues as was described in Section 2. By Theorem 4 the new system stochastically dominates the original one, and by proving stability of it, we clearly establish stability conditions for the original token passing ring. We use induction to establish stability conditions for the nonpersistent queues in the new system, while the stability condition for a persistent queue is established by using Loynes criteria.

To fulfill the above plan, we start by showing a result that will be useful in proving the condition for a persistent queue in the dominant system. More formally, as in Section 2 we consider a doubly infinite sequence of i.i.d random variables $\{S^{n,k}\}_{n,k=1}^{\infty}$ and recall the definition $\tilde{L}^n = \min\{k : \sum_{j=1}^{k} S^{n,j} \geq \tau\}$, $0 \leq \tau < \infty$. We consider further a queue with vacations such that upon the nth arrival of the server to the queue, $\tilde{L}^n$ customers (dummy if necessary) are served and then the server goes for a vacation. The service times of the customers served during the nth visit of the server are the random variables $S^{n,k}$, $k = 1, \cdots, \tilde{L}^n$. Let $\{C^n\}_{n=1}^{\infty}$ be the process of cycle times (time intervals between successive visits to the queue). It is assumed that the processes $\{C^n, \tilde{L}^n\}_{n=1}^{\infty}$ are jointly *stationary and ergodic* (no independence is required). The arrival process A^t to this queue is a Poisson process with parameter λ, independent of the processes $\{C^n, \tilde{L}^n\}_{n=1}^{\infty}$. Let N^n represent the queue length at the beginning of the nth cycle. By X^n we denote the number of customers arrived during the nth cycle. Since A^t is Poisson and independent of the processes $\{C^n, \tilde{L}^n\}_{n=1}^{\infty}$, the processes $\{X^n, \tilde{L}^n\}_{n=1}^{\infty}$ are jointly stationary and ergodic, and $EX = \lambda EC$ where $EC = EC^1$. Clearly, the process of queue lengths at the instants of the visits of the server to the queue satisfies the following recurrence

$$N^{n+1} = \max\{N^n + X^n - \tilde{L}^n \ , \ X^n\} \ , \quad n = 1, 2, \cdots \tag{8}$$

Let $\ell = E\tilde{L}^n$. We prove the following stability result.

Lemma 5. *Consider the queueing system just described. If $\lambda EC < \ell$, then the queue is stable.*

Proof. We apply Loynes' scheme to prove the lemma. We may assume without loss of generality that $X^n, \tilde{L}^n$ is a two-sided stationary process, that is, it is defined for $-\infty < n < \infty$. Note next, that the recursion (8) is such that the RHS of it represents a nondecreasing and left continuous (in N^n) function. Therefore, by Lemma 1 in Loynes [9] we conclude that there exists a stationary sequence $\mathcal{N}^k$ satisfying recursion (8), such that N^n converges in distribution to $\mathcal{N}^1$ provided that $N^1 = 0$. Now, we need to find out when $\mathcal{N}^k$ is honest. Recursion (8) is not quite the same as the one treated by Loynes, however we can use similar arguments as follows.

By telescoping the recurrence (8) we immediately obtain for $n \geq 2$

$$N^n = \max_{1 \leq r \leq n-1} \{X^r + \sum_{k=r+1}^{n-1} \overline{X}^k\} \ . \tag{9}$$

where $\overline{X}^k = X^k - \tilde{L}^k$, provided $N^1 = 0$. Arguing as in Loynes [9] we have that $\mathcal{N}^k$ is honest if and only if

$$\limsup_{r \to \infty} \{X^{-r} + \sum_{k=1}^{r} \overline{X}^{-k}\} < \infty \ . \tag{10}$$

Observe now that

$$X^{-r} + \sum_{k=1}^{r-1} \overline{X}^{-k} = r\left(\frac{\sum_{k=1}^{r} X^{-k}}{r} - \frac{r-1}{r}\frac{\sum_{k=1}^{r-1} \tilde{L}^{-k}}{r-1}\right), \quad r = 2, 3, \cdots$$

Since by the ergodicity of the sequences X^n, $\tilde{L}^n$ we have that $\lim_{r\to\infty} \sum_{k=1}^{r} X^{-k}/r = \lambda EC$ and $\lim_{r\to\infty} \sum_{k=1}^{r} \tilde{L}^{-k}/r = \ell$, the condition $\lambda EC < \ell$ assures the validity of (10). The assumption that $N^1 = 0$ can be removed as in [9]. ∎

Now we are ready to prove our main result described in Proposition 1. In the next theorem we show that the conditions of the Proposition are sufficient.

Theorem 6. *The Markov chain $\mathbf{N}^n(i)$ representing the queue lengths in the token passing ring when it visits queue $i \in \mathcal{M}$ is ergodic if*

$$\lambda_j < \frac{\ell_j}{u_0}(1 - \rho_0) \qquad \text{for all} \quad j \in \mathcal{M} \tag{11}$$

where $\rho_0 = \sum_{j=1}^{M} \rho_j$ and $\ell_i = E\tilde{L}_i$.

Proof. We use mathematical induction. For $M = 1$ the proof is simple since it directly follows from Foster's criterion (cf. [12], [18]).

Now we assume that the theorem is true for $M - 1$ and prove that it can be extended to the M queue case. Let $(\mathcal{U}, \mathcal{S})$, $\mathcal{U} \neq \emptyset$, be a partition of the set $\mathcal{M}$ of M queues into *persistent* and *nonpersistent* queues. Note that the cardinality $|\mathcal{S}|$ of $\mathcal{S}$ is not larger than $M - 1$. Let $\overline{\mathbf{N}}^n(i) = \{\overline{N}_1^n(i), \cdots, \overline{N}_M^n(i)\}$ be the queue lengths when the token visits the ith queue for the nth time in the $(\mathcal{U}, \mathcal{S})$ system in which persistent queues $\mathcal{U}$ send dummy packets as discussed above. Observe that the modified system differs from the original token ring system only in the switchover time from a persistent queue to the successor of that queue in the ring. Specifically, if $i \in \mathcal{U}$, then the switchover times become,

$$\overline{U}_i^k = \Delta_i^k + U_i^k,$$

where Δ_i^k is the time needed to service the dummy messages at node i (if any), and Δ_i^k satisfies condition (A6) of Section 2. Therefore, according to Theorem 4, if $\mathbf{N}^1(1) = \overline{\mathbf{N}}^1(1)$, then

$$\mathbf{N}^n(j) \leq_{st} \overline{\mathbf{N}}^n(j), \quad \text{for all } n, \quad j \in \mathcal{M}. \tag{12}$$

Note now that the queues in $\mathcal{S}$ constitute a token passing ring with $|\mathcal{S}|$ stations satisfying conditions (A1)-(A5) of Section 2, whose operation is independent of the interarrival processes in the persistent queues. The total average switchover time $\overline{u}_0$ to this ring is equal to

$$\overline{u}_0 = u_0 + \sum_{i \in \mathcal{U}} \ell_i s_i .$$

Let the queue lengths in such a system be denoted as $\{\overline{\mathbf{N}}^n_{\mathcal{S}}(i)\}_{i \in \mathcal{S}}$. Clearly, $\overline{\mathbf{N}}^n_{\mathcal{S}}$ is a Markov chain, and since $|\mathcal{S}| \leq M - 1$ we can apply the induction hypothesis. Hence, for $i \in \mathcal{S}$, $\overline{\mathbf{N}}^n_{\mathcal{S}}$ is ergodic if

$$\lambda_i < \frac{\ell_i}{u_0 + \sum_{i \in \mathcal{U}} \ell_i s_i} \left(1 - \sum_{i \in \mathcal{S}} \rho_i\right) \qquad i \in \mathcal{S} . \tag{13}$$

Assume now that (13) holds, and consider a queue in $\mathcal{S}$, say queue 1 and let $C^n_{\mathcal{S}}(1)$ be the process of cycle lengths (successive visits to queue 1). The process $(\mathbf{N}^n_{\mathcal{S}}(1), C^n_{\mathcal{S}}(1))$ is a Markov chain, as easy to notice. Since $\mathbf{N}^n_{\mathcal{S}}(1)$ is ergodic, it is easy to see that also the process $\{\overline{\mathbf{N}}^n_{\mathcal{S}}(1), C^{n-1}_{\mathcal{S}}(1)\}_{n=2}^{\infty}$ is stationary and ergodic (for details see [3]). Now we are in position to use use Lemma 5 which implies that the process $\overline{N}^n_i(1)$, $i \in \mathcal{U}$ is stable provided that

$$\lambda_i < \frac{\ell_i}{EC^1_{\mathcal{S}}(1)} = \frac{\ell_i}{u_0 + \sum_{i \in \mathcal{U}} \ell_i s_i} \left(1 - \sum_{i \in \mathcal{S}} \rho_i\right) \qquad i \in \mathcal{U} , \tag{14}$$

where the equality in (14) follows from the fact that by Theorem 3,

$$EC^1_{\mathcal{S}}(1) = \frac{\overline{u}_0}{1 - \sum_{i \in \mathcal{S}} \rho_i} = \frac{u_0 + \sum_{j \in \mathcal{U}} \ell_j s_j}{1 - \sum_{j \in \mathcal{S}} \rho_j} . \tag{15}$$

Since the process $\overline{N}^n_i(1)$, $i \in \mathcal{S}$ is stable by construction, it follows from (12) that the irreducible, aperiodic Markov chain $\mathbf{N}^n(1)$ is substable and therefore, ergodic. The fact that $\mathbf{N}^n(j)$ is ergodic for all $j \in \mathcal{M}$ follows from Theorem 3.

Putting everything together, from (13) and (14) we finally have that the Markov chain $\mathbf{N}^n(j)$ is ergodic for every $j \in \mathcal{M}$ if

$$\lambda_i < \frac{\ell_i}{u_0 + \sum_{i \in \mathcal{U}} \ell_i s_i} \left(1 - \sum_{i \in \mathcal{S}} \rho_i\right) \qquad i \in \mathcal{M} . \tag{16}$$

Since (16) holds for every partition $\mathcal{P} = (\mathcal{S}, \mathcal{U})$ of the set $\mathcal{M}$ such that $\mathcal{S} \neq \mathcal{M}$, we conclude that the sufficient condition for stability of the system is

$$\mathcal{R} = \bigcup_{\mathcal{S} \subset \mathcal{M}} \mathcal{R}_{\mathcal{S}} , \tag{17}$$

where

$$\mathcal{R}_{\mathcal{S}} = \{\lambda = (\lambda_1, \ldots, \lambda_M) : \text{condition (16) holds}\} . \tag{18}$$

Finally, to complete the proof we need to show that

$$\bigcup_{\mathcal{S}\subset\mathcal{M}} \mathcal{R}_{\mathcal{S}} = \{\lambda = (\lambda_1,\ldots,\lambda_M) : \lambda_i < \frac{\ell_i}{u_0}(1 - \sum_{i=1}^{M}\rho_i) \quad i \in \mathcal{M}\} . \tag{19}$$

This requires only algebraic manipulations which are almost identical to the ones in [3], therefore we omit them here. The interesting reader should be able to reproduce this algebra. ∎

Theorem 6 can be extended to imply the stability of the process $\tilde{\mathbf{N}}(t) = (\tilde{N}_1(t),\cdots,\tilde{N}_M(t))$, where $\tilde{N}_i(t)$ is the queue length at queue i at time t. We just only state the following.

Corollary 7. *The process $\tilde{\mathbf{N}}(t)$ is stable if (11) holds.* ∎

Finally, we show in the next theorem that the conditions of Theorem 6 are also necessary for the ergodicity of the Markov chain $\mathbf{N}^n(i)$, $i \in \mathcal{M}$. This will establish necessary condition for stability of the τ-limited token passing ring, and therefore it completes the proof of Proposition 1.

Theorem 8. *If for some $i \in \mathcal{M}$ the Markov Chain $\mathbf{N}^n(i)$ is ergodic, then $\mathbf{N}^n(j)$ is ergodic for every $i \in \mathcal{M}$. Moreover, $\sum_{j=1}^{M}\rho_j < 1$, and*

$$\lambda_j < \frac{\ell_j}{u_0}(1-\rho_0), \quad j \in \mathcal{M} .$$

Proof. The first assertion follows from Theorem 3 and the remark following that theorem. All cycles in the following will refer to queue 1. For simplicity of notation we omit the queue index from the various variables. Let us define,

C^n: length of the nth cycle.

$C^n(r)$: length of the nth cycle during which r customers from queue 1 were served.

$M^n(r)$: number of cycles in regeneration cycle R^n, (see proof of Theorem 3 for the definition of R^n) during which r customers were served. Clearly,

$$R = \sum_{r=0}^{\infty} M(r), \tag{20}$$

where $M(r) = M^1(r)$ and $R = R^1$.

Since (by the ergodicity of the chain $\mathbf{N}^n(1)$) $ER < \infty$, we have the following formulas for the long run averages:

- average length of a cycle during which r customers were served,

$$EC(r) = \lim_{n\to\infty} \frac{\sum_{k=1}^n C^k(r)}{n} = \frac{E(\sum_{k=1}^{M(r)} C^k(r))}{EM(r)} \ ; \tag{21}$$

- probability (proportion) of cycles during which r customers were transmitted,

$$P(r) = \lim_{n\to\infty} \frac{\sum_{k=1}^n M^k(r)}{n} = \frac{EM(r)}{ER}. \tag{22}$$

Consider now the following system.

System S. Upon arrival of the token to queue 1, the number of customers (from queue 1) that will be served in the next cycle enters system **S**. These customers stay in **S** until the token visits queue 1 for the next time, at which time all customers depart.

Clearly, the number of customers that enter system **S** in the nth cycle is L_1^n. Let A_S^t be the number of customers that arrived in system **S** by time t. Recall the definition of the renewal process $\tilde{C}^n$ in the paragraph before Corollary 7. A_S^t is regenerative with respect to $\tilde{C}^n$, and the ergodicity of $\mathbf{N}^n(1)$ implies by Theorem 3 that $E\tilde{C}^n < \infty$. Hence we have that

$$\lambda_S = \lim_{y\to\infty} \frac{A_S^t}{t} = \frac{\sum_{k=1}^R L_1^k}{\sum_{k=1}^R C^k} = \lambda_1, \tag{23}$$

where the last equality follows from (4). Similarly, taking into account that $E(\sum_{r=1}^\infty rM(r)) = E\left(\sum_{k=1}^R L_1^k\right) < \infty$, we have the following formulas for the long-run average queue size, EN_S, and long-run average waiting time, EW_S, in system **S**.

$$EN_S = \frac{E\left(\sum_{r=1}^\infty r\sum_{k=1}^{M(r)} C^k(r)\right)}{E\left(\sum_{k=1}^R C^k\right)} = \frac{\sum_{r=1}^\infty rE\left(\sum_{k=1}^{M(r)} C^k(r)\right)}{EC\,ER} \tag{24}$$

$$EW_S = \frac{E\left(\sum_{r=1}^\infty r\sum_{k=1}^{M(r)} C^k(r)\right)}{E\left(\sum_{r=1}^\infty rM(r)\right)} = \frac{\sum_{r=1}^\infty rE\left(\sum_{k=1}^{M(r)} C^k(r)\right)}{\sum_{r=1}^\infty rE\left(M(r)\right)} \tag{25}$$

Using (21), (22), we derive from (24), (25),

$$EN_S = \frac{\sum_{r=1}^\infty rP(r)EC(r)}{EC}, \tag{26}$$

$$EW_S = \frac{\sum_{r=1}^\infty rP(r)EC(r)}{\sum_{r=1}^\infty rP(r)}. \tag{27}$$

As in Theorem 1, let L_1 be a random variable distributed as the steady state distribution of the process $\{L_1^n\}_{n=1}^{\infty}$. Then $\sum_{r=1}^{\infty} rP(r) = EL_1$. Since no more than $\tilde{L}_1^n$ customers from queue i are served during the nth cycle, it is easy to see that $L_1 \leq_{st} \tilde{L}_1^1$. If $EL_1 = E\tilde{L}_1^1$, then the stochastic dominance relation implies that $P(L_1 = 0) = P(\tilde{L}_1^1 = 0) = 0$ and from (22) it follows that $EM(0) = 0$. But then, $P(L^n \geq 1, \ \ n = 1, 2\cdots) = 1$ and since $L^n \geq 1$ if and only if $N^n(1) \geq 1$, we have that $P(N^n(1) = 0, \ n = 1, 2\cdots) = 0$ which contradict the ergodicity of the chain $\mathbf{N}^n(1)$. Therefore, $\sum_{r=1}^{\infty} rP(r) < E\tilde{L}_1^1 = \ell_1$ and

$$EW_S > \frac{\sum_{r=1}^{\infty} rP(r)EC(r)}{\ell_1}. \tag{28}$$

Using Little's law (cf. [13]), (23) (26) and (28) we have

$$EN_S = \lambda_S EW_S > \lambda_1 \frac{EN_S EC}{\ell_1},$$

and therefore,

$$\lambda_1 EC < \ell_1, \tag{29}$$

and this proves Theorem 8, and completes the proof of Proposition 1. ■

References

[1] S. Asmussen, *Applied Probability and Queues*, John Wiley&Sons, Chichester 1987.

[2] Chang C., Thomas J., and Kiang S., On the Stability of Open Networks: an Unified Approach by Stochastic Dominance, IBM RC 18343, 1992.

[3] L. Georgiadis and W. Szpankowski, Stability of Token Passing Rings, *Queueing Systems*, 11 (1992) pp. 7-33.

[4] O. Ibe and X. Cheng, Stability Conditions for Multiqueue Systems with Cycle Service, *IEEE Trans. Automatic Control*, 33 (1988) 102-103.

[5] P. Kuehn, Multiqueue Systems with Nonexhaustive Cycle Service, *The bell System Technical Journal*, 58 (1979) 671-698.

[6] K.K. Leung and M. Eisenberg, A Single-Server Queue with Vacations and Gated Time-Limited Service, *IEEE Trans. Commun.*, 38 (1990) 1454-1462.

[7] K.K. Leung and M. Eisenberg, A Single-Server Queue with Vacations and Non-Gated Time-Limited Service, *Perf. Evaluation J.*, 12 (1991) 115-125.

[8] H. Levy, M. Sidi and O. Boxma, Dominance Relations in Polling Systems, *Queueing Systems*, 6 (1990) 155-172.

[9] R. Loynes, The stability of a queue with non-independent inter-arrival and service times, *Proc. Camb. Philos. Soc.*, 58 (1962), 497–520.

[10] V.A. Malyshev and M.V Mensikov, Ergodicity, continuity and analyticity of countable Markov chains, *Trans. Moscow Math. Soc.* (1981), 1-18.

[11] S. Meyn and D. Down, Stability of Generalized Jackson Networks, *Ann. Appl. Probab.*, to appear.

[12] S. Meyn and R.L. Tweedie, Criteria for Stability of Markovian Processes I: Discrete Time Chains, *Adv. Appl. Prob.*, 24, 542-574, 1992.

[13] S. Stidham, Jr., A last word on $L = \lambda W$. *Oper. Res.*, 22 (1974) , pp.417-421.

[14] W. Szpankowski, Towards computable stability criteria for some multidimensional stochastic processes, in *Stochastic Analysis of Computer and Communication Systems* (ed. H. Takagi), Elsevier Science Publications B. V. (North-Holland), (1990), pp. 131-172.

[15] W. Szpankowski, Stability Conditions for Some Distributed Systems: Buffered Random Access Systems, *Adv. Appl. Probab.*, to appear.

[16] H. Takagi, *Analysis of Polling Systems*, The MIT Press, Cambridge, Mass. 1988.

[17] H. Takagi, Queueing Analysis of Polling Models, *ACM Computing Surveys*, 20 (1988) 5-28.

[18] R. L. Tweedie, Criteria for classifying general Markov chains, *Adv. Appl. Prob.*, 8 (1976), 737–771.

[19] J. Walrand, *An Introduction to Queueing Networks*, Prentice Hall, New Jersey 1988.

[20] K. Watson, Performance Evaluation of Cyclic Service Strategies – A Survey. *Proc. PERFORMANCE'84*, E. Gelenbe Ed. (Paris 1984).

Unité de Recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)

Unité de Recherche INRIA Lorraine Technopôle de Nancy-Brabois - Campus Scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 VILLERS LES NANCY Cedex (France)
Unité de Recherche INRIA Rennes IRISA, Campus Universitaire de Beaulieu 35042 RENNES Cedex (France)
Unité de Recherche INRIA Rhône-Alpes 46, avenue Félix Viallet - 38031 GRENOBLE Cedex (France)
Unité de Recherche INRIA Sophia Antipolis 2004, route des Lucioles - B.P. 93 - 06902 SOPHIA ANTIPOLIS Cedex (France)

EDITEUR
INRIA - Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)

ISSN 0249 - 6399

RR.1996